DOCUMENT RESUME

ED 363 659 TM 020 781

AUTHOR Kuo, Lynn; Cohen, Michael P.
TITLE Bayesian Analysis for Linearized Multi-Stage Models in Quantal Bioassay.
PUB DATE [93]
NOTE 23p.; Author is affiliated with the National Center for Education Statistics (NCES).
PUB TYPE Statistical Data (110) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Bayesian Statistics; *Equations (Mathematics); *Estimation (Mathematics); *Mathematical Models; *Predictive Measurement; Sampling; Selection; Statistical Inference; Tables (Data)
IDENTIFIERS Conditional Predicted Ordinates; Cross Validation; Dose Response Curve; Gibbs Sampling; *Linear Models; Multistage Models; *Quantal Bioassay; Risk Assessment

ABSTRACT

Bayesian methods for estimating dose response curves in quantal bioassay are studied. A linearized multi-stage model is assumed for the shape of the curves. A Gibbs sampling approach with data augmentation is employed to compute the Bayes estimates. In addition, estimation of the "relative additional risk" and the "risk specific dose" is studied. Model selection based on conditional predictive ordinates from cross-validated data is developed. Contains 22 references and 8 tables.) (Author)

Bayesian Analysis for Linearized Multi-Stage Models in Quantal Bioassay

LYNN KUO and MICHAEL P. COHEN *

Abstract

Bayesian methods for estimating dose response curves in quantal bioassay are studied. A linearized multi-stage model is assumed for the shape of the curves. A Gibbs sampling approach with data augmentation is employed to compute the Bayes estimates. In addition, estimation of the "relative additional risk" and the "risk specific dose" is studied. Model selection based on conditional predictive ordinates from cross-validated data is developed.

KEY WORDS: Dose response curve; Gibbs sampling; Model selection; Risk; Stochastic substitution.

*Lynn Kuo is Associate Professor, Department of Statistics, University of Connecticut, Storrs, CT 06269. Michael P. Cohen is Mathematical Statistician, Statistical Standards and Methodology Division, National Center for Education Statistics, Washington, DC 20208-5654. The views are those of the authors, and no official support by the U.S. Department of Education is intended or should be inferred. The authors thank Herbert Lacayo for acquainting them with this problem.

1. Introduction

Several models are available to explain the relationship between the response probability and the dosage in quantal bioassay. These include the probit, logit, one-hit, multi-hit, multistage, and Weibull models. Krewski and Van Ryzin (1981) provide a review. The maximum likelihood method is usually employed to estimate the response curves. The Bayesian approach to these problems has been limited to a few simplified cases. Petrasovitas and Cornell (1975) propose approximations to the Bayes estimates for the one-hit model with single-dose assay. Zellner and Rossi (1982) employ numerical integration techniques for computing Bayes estimates in probit and logit models with noninformative priors. Recently, some advances have been made for more complex situations. Albert and Chib (1991) develop a Gibbs sampling approach for the probit model. They also extend their methods to polytomous responses and the t-link distribution. Gelfand and Kuo (1991) develop Gibbs sampling approaches for nonparametric dose response curves with Dirichlet process priors and with product of beta priors. They also extend their models and methods to polytomous responses. Ramgopal, Laud, and Smith (1992) extend the work of Gelfand and Kuo to convex, concave, or ogive potency curves. Nagaraj (1991) proposes an importance sampling approach for the linearized multi-stage model.

The linearized multi-stage model specifies the probability of a subject's positive response to a dose at d as

$$P(d|\beta) = 1 - e^{-(\beta_0+\beta_1 d+\cdots+\beta_k d^k)} \tag{1.1}$$

with $\beta_j \geq 0,\ j = 0,\ldots,k$. This model is currently used by the U.S. Environmental Protection Agency (EPA) for risk assessment. As discussed

by Crump, Hoel, Langley, and Peto (1976) and Crump (1979), the model is based on the assumption that induction of toxic effects results from a number of different random biological events, with the age-specific rate of occurrence of each event linearly related to dose. When $k = 1$, this is the one-hit model described by Armitage and Doll (1961). The model in (1.1) can also be considered as a generalized linear model (McCullagh and Nelder 1989) with complementary log as the link function.

To apply Bayesian analysis for the unknown parameters $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)$, we assume that the $\beta_0, \ldots, \beta_k$ are independent, where β_j has a gamma distribution $\Gamma(\alpha_j, \gamma_j)$ with mean α_j/γ_j for $j = 0, \ldots, k$. The gamma distributions are chosen partly for convenience. The family of two parameter gamma densities, though, incorporates densities with both increasing and decreasing failure rates. Often statisticians have some prior belief about the location and spread of the distribution of β_j for each j. Let μ_j and σ_j denote a prior guess of this location and spread. Then α_j and γ_j can be chosen by the method of matching moments so that they satisfy $\mu_j = \alpha_j/\gamma_j$ and $\sigma_j^2 = \alpha_j/\gamma_j^2$.

In risk assessment, subjects are divided into groups, say I groups, where for $i = 1, \ldots, I$, n_i subjects are given the substance at dosage level d_i. The number of casualties observed in group i is denoted by x_i. Let $\mathbf{x} = (x_1, \ldots, x_I)$. Given $\boldsymbol{\beta}$, the subjects' responses are assumed to be independent. Therefore the likelihood function is

$$L(\boldsymbol{\beta};\, \mathbf{x}) = \prod_{i=1}^{I} \binom{n_i}{x_i} P(d_i|\boldsymbol{\beta})^{x_i} (1 - P(d_i|\boldsymbol{\beta}))^{n_i - x_i}.$$

It can be shown that the posterior distribution is a mixture of products of gamma distributions. This mixture increases in complexity as I

increases. This paper provides a Gibbs sampling approach to computing the Bayes estimates. It can be used for any number of groups. This approach augments the data by latent variables that convert the integrals (cumulative probabilities and survival probabilities) in the likelihood to probability density functions. This augmentation facilitates us in specifying the conditional density of β given the latent variables. A repeated sampling scheme that generates variates from this conditional density and the conditional density of the latent variables given β and the data allows us to approximate the posterior distribution.

The dose response model in (1.1) is often used for low-dose extrapolation, where the potency of a toxic substance on human beings (at low dose) is inferred from experiments conducted at much higher doses. Practitioners are often interested in "Relative Additional Risk" (RAR) and "Risk Specific Dose" (RSD). We provide methods of Bayesian inference for these quantities. RAR can be interpreted as the conditional probability of a subject acquiring cancer, say, at dose level d, given that no cancer has been developed in the absence of the carcinogen. RSD measures the dosage such that the probability of a casualty is no more than a given number.

Because the model in (1.1) depends on k, the number of parameters, it is natural to ask how to select k. One route to selecting k has been for biologists to propose theory to explain the model. Another route, more empirical, is to fit the models to data. Chi-squared goodness of fit tests are usually employed by frequentists for model checking. A criterion for model selection based on Bayesian predictive ordinates from cross-validated data is adopted in this paper.

Section 2 describes the Gibbs sampling approach. Section 3 addresses

estimation of RAR and RSD. Section 4 develops the criterion for model selection. Examples are given in Section 5, and some concluding remarks are provided in Section 6.

2. Gibbs Sampling

A Gibbs sampling approach specific for our problem is developed here to compute the posterior distribution of β. One can refer to Geman and Geman (1984), Tanner and Wong (1987), and Gelfand and Smith (1990) for general discussion of Gibbs sampling.

Given the likelihood and the prior, we observe that the posterior density of β is

$$f(\beta|\mathbf{x}) \propto \prod_{i=1}^{I} \left(1 - e^{-(\beta_0+\beta_1 d_i+\cdots+\beta_k d_i^k)}\right)^{x_i} \prod_{j=0}^{k} \beta_j^{\alpha_j-1} e^{-[\gamma_j+\sum_{i=1}^{I}(n_i-x_i)d_i^j]\beta_j}. \quad (2.1)$$

To sample β from (2.1) by stochastic substitution with augmented data $\mathbf{z}$, we can sample repeatedly from $f(\mathbf{z}|\beta,\mathbf{x})$, the conditional distribution of $\mathbf{z}$ given β and $\mathbf{x}$, and $f(\beta|\mathbf{z},\mathbf{x})$, the conditional distribution of β given $\mathbf{z}$ and $\mathbf{x}$. Therefore, it is desirable to augment the data by the latent variables $\mathbf{z}$, such that we can easily sample variates from the two conditional densities $f(\mathbf{z}|\beta,\mathbf{x})$ and $f(\beta|\mathbf{z},\mathbf{x})$. For each i, let us consider the latent variables $\{z_{ij}\}$, $j = 1,\ldots,x_i$, as i.i.d. random variables with an exponential distribution with mean $1/(\beta_0 + \beta_1 d_i + \ldots + \beta_k d_i^k)$. Observe that

$$P(z_{ij} < 1) = 1 - e^{-(\beta_0+\beta_1 d_i+\cdots+\beta_k d_i^k)}.$$

Assume the distribution of the $\mathbf{z}_i = (z_{i1},\ldots,z_{i,x_i})$ are independent for $i = 1,\ldots,I$. Let $\mathbf{z} = (\mathbf{z}_1,\ldots,\mathbf{z}_I)$. Then the conditional density of $\mathbf{z}$ given β

and $\mathbf{x}$ is

$$g(\mathbf{z}|\boldsymbol{\beta},\mathbf{x}) = \prod_{i=1}^{I}\prod_{j=1}^{x_i} g_i(z_{ij}|\boldsymbol{\beta},\mathbf{x}), \tag{2.2}$$

where g_i is derived from an exponential distribution with mean $1/(\beta_0 + \beta_1 d_i + \cdots + \beta_k d_i^k)$ by truncation on the right by 1. To obtain the conditional density of $\boldsymbol{\beta}$ given $\mathbf{z}$ and $\mathbf{x}$, set $s_i(\boldsymbol{\beta}) = \beta_0 + \beta_1 d_i + \cdots + \beta_k d_i^k$ and observe

$$\begin{aligned} f(\boldsymbol{\beta}|\mathbf{z},\mathbf{x}) \;\propto\; & \prod_{j=0}^{k} \beta_j^{\alpha_j - 1} e^{-[\gamma_j + \sum_{i=1}^{I}(n_i - x_i)d_i^j]\beta_j} \\ & \times \prod_{i=1}^{I} [s_i(\boldsymbol{\beta})]^{x_i} e^{-s_i(\boldsymbol{\beta})(\sum_{j=1}^{x_i} z_{ij})}. \end{aligned} \tag{2.3}$$

This suggests another data augmentation step is needed to facilitate the generation of the $\boldsymbol{\beta}$ variate given the latent variables. For each i, consider $\mathbf{w}_i = (w_{i0}, \ldots, w_{ik})$ distributed as a multinomial distribution with parameters x_i and cell probabilities $\left(\frac{\beta_0}{s_i(\boldsymbol{\beta})}, \frac{\beta_1 d_i}{s_i(\boldsymbol{\beta})}, \ldots, \frac{\beta_k d_i^k}{s_i(\boldsymbol{\beta})} \right)$. Moreover, generate the $\mathbf{w}_1, \ldots, \mathbf{w}_I$ independently. Let $\mathbf{w}$ denote $(\mathbf{w}_1, \ldots, \mathbf{w}_I)$. Then the joint density of the $\mathbf{w}$ is given by

$$\begin{aligned} f(\mathbf{w}|\boldsymbol{\beta},\mathbf{x}) &= \prod_{i=1}^{I} h_i(\mathbf{w}_i|\mathbf{x},\boldsymbol{\beta}) \\ &= \prod_{i=1}^{I} \binom{x_i}{w_{i0}\cdots w_{ik}} \left(\frac{\beta_0}{s_i(\boldsymbol{\beta})}\right)^{w_{i0}} \left(\frac{\beta_1 d_i}{s_i(\boldsymbol{\beta})}\right)^{w_{i1}} \cdots \left(\frac{\beta_k d_i^k}{s_i(\boldsymbol{\beta})}\right)^{w_{ik}}. \end{aligned} \tag{2.4}$$

Let $z_{i+} = \sum_{j=1}^{x_i} z_{ij}$. Now the conditional density of $\boldsymbol{\beta}$ given $\mathbf{z}, \mathbf{w}$, and $\mathbf{x}$ can be verified to be the product of independent gamma densities with parameters $\hat{\alpha}_j = \sum_{i=1}^{I} w_{ij}$ and $\hat{\gamma}_j = \sum_{i=1}^{I} d_i^j (z_{i+} + n_i - x_i)$,

$$f(\boldsymbol{\beta}|\mathbf{z},\mathbf{w},\mathbf{x}) \propto \prod_{j=0}^{k} \beta_j^{\hat{\alpha}_j - 1} e^{-\hat{\gamma}_j \beta_j}. \tag{2.5}$$

Having explained the conditional densities needed in the substitution algorithm, we list the steps used to generate the samplers.

(1) Start with an initial choice of the β, denoted by $\beta^{(0)}$. Set the counter in the superscript to 0.

(2) Generate $\mathbf{z}^{(1)}$ and $\mathbf{w}^{(1)}$ independently from $f(\mathbf{z}|\beta^{(0)}, \mathbf{x})$ as in (2.2) and $f(\mathbf{w}|\beta^{(0)}, \mathbf{x})$ as in (2.4).

(3) Generate $\beta^{(1)}$ from $f(\beta|\mathbf{z}^{(1)}, \mathbf{w}^{(1)}, \mathbf{x})$ as in (2.5).

(4) Repeat steps *(1)*, *(2)*, and *(3)* with the updated counter in all the superscripts, that is, with $\beta^{(l+1)}$ replacing $\beta^{(l)}$, etc., in each step.

The generation of z_{ij} in *(3)* from the truncated density can be facilitated by use of the equation $z_{ij} = G^{-1}(G(0)+U_{ij}(G(1)-G(0)))$, where the U_{ij} are independent uniform [0,1] variates, and G is the exponential distribution function with mean $1/s_i(\beta)$. Then, $z_{ij} = -\ln(1 - U_{ij}(1 - e^{-s_i(\beta)}))/s_i(\beta)$, independently for $j = 1, \ldots, x_i$.

Suppose we also replicate the above iterations R times by using independent initial choices of the β. Let the β, $\mathbf{z}$, and $\mathbf{w}$ variates generated in the L^{th} step of the iteration for the r^{th} replication be denoted by $\beta_r^{(L)}, \mathbf{z}_r^{(L)}$, and $\mathbf{w}_r^{(L)}$, for $r = 1, \ldots, R$. Similar notation is applied to the variates w_{ij}, z_{ij}, z_{i+}, etc. Then the posterior distribution of β is estimated by

$$\hat{f}(\beta|\mathbf{x}) \approx \frac{1}{R} \sum_{r=1}^{R} f(\beta|\mathbf{z}_r^{(L)}, \mathbf{w}_r^{(L)}, \mathbf{x}), \tag{2.6}$$

where f on the right is the product of gamma distributions described as in (2.5). Since the β_j, $j = 0, \ldots, k$, are independent in (2.5), the posterior mean of β_j can be approximated by

$$\hat{\beta}_j = \frac{1}{R} \sum_{r=1}^{R} \frac{\alpha_j + \sum_{i=1}^{I} w_{ij,r}^{(L)}}{\gamma_j + \sum_{i=1}^{I} d_i^j (z_{i+,r}^{(L)} + n_i - x_i)}. \tag{2.7}$$

The computations in (2.6) and (2.7) use only the variates generated in the last step of each iteration. A more efficient algorithm has been proposed by Gelman and Rubin (1992) where estimates are produced by averaging over all variates generated in the second halves of the iterations for all replications. They also study the choices of L and R by analysis of variance techniques.

We will use the empirical values $\{\beta_{j,r}^{(L)}\}_{r=1}^{R}$ from the replicated samples to construct the predictive interval for β_j.

3. Relative Additional Risk and Risk Specific Dose

The computation of some functionals of the parameters by stochastic substitution is discussed in this section.

The relative additional risk over the background at a dose of level d is defined by

$$A(d) = \frac{P(d) - P(0)}{1 - P(0)} = 1 - e^{\beta_1 d + \cdots + \beta_k d^k}.$$

The relative additional risk can be interpreted as the conditional probability of acquiring cancer, for example, at dose level d given that no cancer has developed in the absence of the carcinogen. Observe that $A(d) \approx \beta_1 d$ at low dose. Let $\hat{A}_r^{(L)}(d)$ denote an estimate of $A(d)$ computed from the L^{th} iteration and the r^{th} replication in the sampling procedure. Then $\hat{A}_r^{(L)}(d)$ can be computed from (2.5) and the moment generating functions of the

gamma distributions

$$\hat{A}_r^{(L)}(d) \approx 1 - \prod_{j=0}^{k} \left(\frac{\gamma_j + \sum_{i=1}^{I} d_i^j (n_i - x_i + z_{i+,r}^{(L)})}{d^j + \gamma_j + \sum_{i=1}^{I} d_i^j (n_i - x_i + z_{i+,r}^{(L)})} \right)^{\alpha_j + \sum_{i=1}^{I} w_{ij,r}^{(L)}}. \quad (3.1)$$

The Bayes estimate for $A(d)$, denoted by $\hat{A}(d)$, is computed from

$$\hat{A}(d) = \frac{1}{R} \sum_{r=1}^{R} \hat{A}_r^{(L)}(d). \quad (3.2)$$

The predictive interval for $A(d)$ can be computed from the empirical measure that assigns weight $1/R$ to each $\hat{A}_r^{(L)}(d)$ for $r = 1, \ldots, R$.

The Risk Specific Dose (RSD) is defined as the dose level such that the probability of a casualty $P(d)$ is at most θ. $RSD(\theta)$ is the value d such that $P(d) = \theta$. Equivalently, $RSD(\theta)$ is defined to be the largest real root of the following equation:

$$\ln(1 - \theta) + \beta_0 + \beta_1 d + \cdots + \beta_k d^k = 0. \quad (3.3)$$

Since the parameters $\{\beta_j\}_{j=0}^{k}$ are unknown, we propose to estimate the β_j, $j = 0, \ldots, k$, by (2.6) first, then find the roots of (3.3) where β_j is replaced by $\hat{\beta}_j$, the estimates of β_j, for all $j = 0, \ldots, k$. Observe that when $k = 1$, then $\widehat{RSD}(\theta) = -[\hat{\beta}_0 + \ln(1 - \theta)]/\hat{\beta}_1$.

4. Model Selection

The selection of k, the degree of the polynomial in the exponent of (1.1), is studied in this section. A criterion based on the conditional predicted ordinates from cross-validated data is adopted for model selection. This approach, called pseudo-marginal likelihood, was first proposed by Geisser

and Eddy (1979). The role of using predictive densities from cross-validated data for model determination was also supported by Stone (1974), Box (1980), and Pettit and Young (1990). Box (1980) states that the predictive distribution enables one to criticize the entertained model in the light of current data. Gelfand, Dey, and Chang (1992) have provided more detailed discussion of Bayesian model selection and its implementation by means of Gibbs sampling.

Because the selection criterion was developed for independent observations, we need to reconsider our data from this viewpoint. Suppose each subject is given a dose independently. The response of each subject is either positive or negative. Therefore, we observe n independent Bernoulli random variables with probability of success p_s, $s = 1, \ldots, n$. If $d_{i(s)}$ is the dosage for subject s, then

$$p_s = 1 - e^{-(\beta_0 + \beta_1 d_{i(s)} + \cdots + \beta_k d_{i(s)}^k)}.$$

Let $\mathbf{y}_{(s)}$ denote the whole data set with the s^{th} observation y_s deleted. The conditional predicted ordinate (CPO) for y_s, denoted by c_s, is defined by

$$c_s = f(y_s|\mathbf{y}_{(s)}) = \int f(y_s|\beta, \mathbf{y}_{(s)}) f(\beta|\mathbf{y}_{(s)}) d\beta. \tag{4.1}$$

Note c_s depends on k. We suppress this dependence in the notation for the time being for simplicity.

The estimated density $\hat{f}(\beta|\mathbf{y}_{(s)})$ can be computed as in (2.6) with y_s deleted from the data set. Let $\hat{\alpha}_{j,r,(s)}^{(L)}$ and $\hat{\gamma}_{j,r,(s)}^{(L)}$ denote the parameters of the gamma density of β_j in (2.5) and (2.6) computed from the data with y_s deleted. Since $f(y_s|\beta, \mathbf{y}_{(s)}) = f(y_s|\beta) = p_s^{y_s}(1 - p_s)^{1-y_s}$, the computation of (4.1) is straightforward with the use of the moment generating function

of a gamma distribution. If y_s is 0 and the s^{th} subject is given the dosage d, then c_s is approximated by

$$\hat{c}_s \approx \frac{1}{R}\sum_{r=1}^{R}\prod_{j=0}^{k}\left(\frac{\hat{\gamma}_{j,r,(s)}^{(L)}}{\hat{\gamma}_{j,r,(s)}^{(L)}+d^j}\right)^{\hat{\alpha}_{j,r,(s)}^{(L)}}. \tag{4.2}$$

If y_s is 1 and the s^{th} subject is given the dosage d, then c_s is approximated by one minus the right side of(4.2). The variance of CPO c_s is given by

$$V(c_s) = \int e^{-2(\beta_0+\beta_1 d+\cdots+\beta_k d^k)} f(\beta|\mathbf{y}_{(s)})d\beta - \left(\int e^{-(\beta_0+\beta_1 d+\cdots+\beta_k d^k)} f(\beta|\mathbf{y}_{(s)})d\beta\right)^2.$$

It is approximated by

$$\hat{V}(c_s) = \frac{1}{R}\sum_{r=1}^{R}\prod_{j=0}^{k}\left(\frac{\hat{\gamma}_{j,r,(s)}^{(L)}}{\hat{\gamma}_{j,r,(s)}^{(L)}+2d^j}\right)^{\hat{\alpha}_{j,r,(s)}^{(L)}} - \hat{c}_s^{\,2}. \tag{4.3}$$

Since c_s depends on k, we now use the notation $c_s(k)$ for c_s based on the model with a k degree polynomial in the generalized linear model. Define the pseudo-marginal likelihood $C(k) = \prod_{s=1}^{n} \hat{c}_s(k)$. We select the k with the largest pseudo-marginal likelihood. Equivalently, we select the model with the largest $\sum_{s=1}^{n} \ln\ \hat{c}_s(k)$.

5. Numerical Examples

5.1 Bayes Inference

The data set is taken from Howe, Crump, and Van Landingham (1986). Nagaraj (1991) also used it for illustration. There are 200 animals divided into four groups with 50 animals in each group. The dosages assigned to the

groups are 0, 1, 2, and 3. The observed number of positive responses x in each group is $(0, 2, 10, 30)$. Tables 1–3 exhibit the various Bayes estimates computed based on the Gibbs sampling method with 50 iterations and 1000 replications. Three models with $k = 1, 2, 3$ are considered, each with three choices of priors. The choices are (1) $\alpha_j = 2$ and $\gamma_j = 20$ for all j, (2) $\alpha_j = \gamma_j = 1$ for all j, and (3) $\alpha_j = 1$ and $\gamma_j = .01$. The priors are chosen to center at .1, 1, and 100 with small, medium, and large variances. The parameters and functionals of interest are β_j, $j = 0, \ldots, k$, and the Relative Additional Risk (RAR) $A(d)$ evaluated at $d = .5, 1, 2$, and 3 ($A(0) = 0$ by definition). For each of these functionals, we display the point estimate (posterior mean), the posterior standard deviation, and the posterior 95% confidence interval. The Bayes estimates for $P(d)$ evaluated at the dosage levels are also given. The corresponding standard deviations and confidence intervals can be computed from the Gibbs samplers by methods similar to those for β_j and $A(d)$. They are omitted here for brevity. The risk specific dose (RSD) for $\theta = .025, .05, .075, .1, .125$, and .15 is provided.

The results of these tables show that the Bayes estimates for the second and third choices of the priors are very close to each other. The effect of the small mean and variance of the first prior can be seen from the highest coefficient (β_k) for each model. As k increases, the Bayes estimates of $P(d)$ for $d = 0, 1, 2$, and 3 approach the nonparametric maximum likelihood estimates, which are 0, .04, .2, and .6, respectively.

5.2 Model Selection

The numerical results for the model selection are given here. We have considered each y_s, $s = 1, \ldots, 200$, to have a Bernoulli distribution with

the probability of a success (positive response) to be $P(d)$, where d is the dose given to the subject s. Given the data $\mathbf{x} = (0, 2, 10, 30)$ observed at the dosages 0, 1, 2, and 3, we can consider $y_s = 0$ for s = 1–50, 53–100, 111–150, 181–200, and $y_s = 1$ for the rest of the indices.

Because of the grouped data $\mathbf{x}$, many of the CPO's will have the same value. In fact, instead of computing 200 CPO's, we need only compute 7 of them. Let $\mathbf{n}$ denote the number of subjects in each group. Then the computation of the CPO for $s = 1, \ldots, 50$ is based on $\mathbf{n} = (49, 50, 50, 50)$ and $\mathbf{x} = (0, 2, 10, 30)$; the CPO for $s = 51, 52$ is based on $\mathbf{n} = (50, 49, 50, 50)$ and $\mathbf{x} = (0, 1, 10, 30)$, etc.

The values of the CPO's for each prior choice with different models are given in Tables 4–6. Usually we prefer to select a diffuse prior for model selection and let the data do the talking. In our examples, the analysis is relatively insensitive to the prior choice. They all suggest that the model with $k = 2$ improves upon the model with $k = 1$. Although the pseudo-marginal likelihood factor for the model with $k = 3$ is slightly larger than that of the smaller model, the improvement is very marginal. This suggests either model, $k = 2$ or $k = 3$, may be acceptable for the data set. In fact, Table 7, which provides the Gibbs approximation to the pseudo-marginal likelihood for $k = 1, 2, \ldots, 9$, shows that the pseudo-marginal likelihood is essentially flat for $2 \leq k \leq 9$. Table 8 indicates that the value of the coefficient β_1, important for low-dose extrapolation, is not sensitive to the model for $k \geq 3$.

Table 1: Gibbs Approximation to the Bayes Estimates when $k = 1$

	$\alpha_j = 2, \gamma_j = 20$		$\alpha_j = 1, \gamma_j = 1$		$\alpha_j = 1, \gamma_j = .01$	
$\hat{\beta}_0$ (SD)	.022	(.011)	.011	(.008)	.012	(.008)
95% C.I.	(.010,	.050)	(.005,	.034)	(.006,	.034)
$\hat{\beta}_1$ (SD)	.161	(.009)	.175	(.007)	.175	(.007)
95% C.I.	(.138,	.174)	(.157,	.185)	(.157,	.186)
$\hat{A}(.5)$ (SD)	.077	(.015)	.084	(.015)	.084	(.016)
95% C.I.	(.066,	.083)	(.075,	.088)	(.075,	.089)
$\hat{A}(1)$ (SD)	.148	(.024)	.160	(.025)	.160	(.025)
95% C.I.	(.129,	.159)	(.145,	.169)	(.145,	.169)
$\hat{A}(2)$ (SD)	.274	(.039)	.294	(.040)	.294	(.040)
95% C.I.	(.240,	.293)	(.268,	.309)	(.268,	.309)
$\hat{A}(3)$ (SD)	.380	(.049)	.406	(.050)	.406	(.050)
95% C.I.	(.337,	.405)	(.373,	.425)	(.373,	.426)
$\hat{P}(0)$, $\hat{P}(1)$	.021	.167	.011	.170	.012	.170
$\hat{P}(2)$, $\hat{P}(3)$	.290	.395	.302	.414	.304,	.415
$\widehat{RSD}(.025), \widehat{RSD}(.05)$	.022	.184	.079	.228	.076	.225
$\widehat{RSD}(.075), \widehat{RSD}(.10)$	.350	.521	.381	.538	.377	.534
$\widehat{RSD}(.125), \widehat{RSD}(.15)$	.697	.877	.699	.865	.695	.861

Table 2: Gibbs Approximation to the Bayes Estimates when $k = 2$

	$\alpha_j = 2, \gamma_j = 20$		$\alpha_j = 1, \gamma_j = 1$		$\alpha_j = 1, \gamma_j = .01$	
$\hat{\beta}_0$ (SD)	.019	(.009)	.011	(.008)	.011	(.008)
95% C.I.	(.010,	.041)	(.006,	.033)	(.006,	.034)
$\hat{\beta}_1$ (SD)	.037	(.021)	.029	(.026)	.032	(.026)
95% C.I.	(.008,	.088)	(.004,	.100)	(.004,	.100)
$\hat{\beta}_2$ (SD)	.063	(.011)	.068	(.012)	.067	(.012)
95% C.I.	(.039,	.079)	(.035,	.083)	(.035,	.083)
$\hat{A}(.5)$ (SD)	.034	(.014)	.031	(.015)	.032	(.015)
95% C.I.	(.023,	.053)	(.021,	.058)	(.021,	.057)
$\hat{A}(1)$ (SD)	.096	(.020)	.093	(.021)	.094	(.022)
95% C.I.	(.079,	.121)	(.076,	.125)	(.077,	.128)
$\hat{A}(2)$ (SD)	.279	(.038)	.281	(.038)	.282	(.038)
95% C.I.	(.251,	.296)	(.258,	.299)	(.257,	.300)
$\hat{A}(3)$ (SD)	.491	(.057)	.502	(.057)	.501	(.057)
95% C.I.	(.440,	.527)	(.449,	.535)	(.449,	.532)
$\hat{P}(0)$, $\hat{P}(1)$	.019	.113	.011	.103	.011	.105
$\hat{P}(2)$, $\hat{P}(3)$	.293	.504	.290	.510	.291	.510
$\widehat{RSD}(.025), \widehat{RSD}(.05)$	.139	.479	.288	.582	.275	.568
$\widehat{RSD}(.075), \widehat{RSD}(.10)$	.715	.910	.798	.980	.785	.968
$\widehat{RSD}(.125), \widehat{RSD}(.15)$	1.083	1.240	1.142	1.290	1.131	1.280

Table 3: Gibbs Approximation to the Bayes Estimates when $k = 3$

	$\alpha_j = 2, \gamma_j = 20$		$\alpha_j = 1, \gamma_j = 1$		$\alpha_j = 1, \gamma_j = .01$	
$\hat{\beta}_0$ (SD)	.018	(.008)	.012	(.008)	.011	(.008)
95% C.I.	(.010,	.041)	(.006,	.034)	(.006,	.033)
$\hat{\beta}_1$ (SD)	.030	(.017)	.025	(.019)	.025	(.020)
95% C.I.	(.008,	.071)	(.004,	.074)	(.004,	.073)
$\hat{\beta}_2$ (SD)	.025	(.014)	.023	(.018)	.023	(.018)
95% C.I.	(.005,	.057)	(.002,	.066)	(.002,	.067)
$\hat{\beta}_3$ (SD)	.019	(.006)	.020	(.007)	.020	(.008)
95% C.I.	(.007,	.029)	(.003,	.031)	(.003,	.031)
$\hat{A}(.5)$ (SD)	.023	(.014)	.021	(.014)	.021	(.014)
95% C.I.	(.012,	.041)	(.008,	.043)	(.007,	.041)
$\hat{A}(1)$ (SD)	.071	(.020)	.066	(.022)	.066	(.022)
95% C.I.	(.049,	.100)	(.041,	.101)	(.040,	.100)
$\hat{A}(2)$ (SD)	.268	(.037)	.260	(.038)	.260	(.038)
95% C.I.	(.239,	.291)	(.231,	.288)	(.232,	.289)
$\hat{A}(3)$ (SD)	.560	(.059)	.555	(.061)	.558	(.061)
95% C.I.	(.503,	.604)	(.500,	.601)	(.502,	.601)
$\hat{P}(0)$, $\hat{P}(1)$	.018	.088	.012	.077	.011	.076
$\hat{P}(2)$, $\hat{P}(3)$	.282	.572	.269	.565	.270	.567
$\widehat{RSD}(.025), \widehat{RSD}(.05)$	.195	.624	.370	.741	.381	.748
$\widehat{RSD}(.075), \widehat{RSD}(.10)$	.888	1.091	.983	1.173	.989	1.177
$\widehat{RSD}(.125), \widehat{RSD}(.15)$	1.258	1.405	1.332	1.472	1.335	1.474

Table 4: Gibbs Approximation to the CPO for the Prior with $\alpha_j = 2$ and $\gamma_j = 20$ for all j

	s	1-50	51-52	53-100	101-110	111-150	151-180	181-200	$\ln(\prod \hat{c}_s)$
$k = 1$	$\hat{c}_s$	.980	.163	.833	.285	.708	.389	.601	-78.242
	$\widehat{SD}(c_s)$	(.017)	(.024)	(.024)	(.037)	(.037)	(.047)	(.047)	
$k = 2$	$\hat{c}_s$	.982	.110	.887	.288	.706	.495	.494	-72.743
	$\widehat{SD}(c_s)$	(.015)	(.022)	(.022)	(.037)	(.037)	(.055)	(.055)	
$k = 3$	$\hat{c}_s$	.982	.084	.912	.277	.717	.563	.424	-70.863
	$\widehat{SD}(c_s)$	(.016)	(.021)	(.022)	(.036)	(.036)	(.059)	(.058)	

Table 5: Gibbs Approximation to the CPO for the Prior with $\alpha_j = 1$ and $\gamma_j = 1$ for all j

	s	1-50	51-52	53-100	101-110	111-150	151-180	181-200	$\ln(\prod \hat{c}_s)$
$k = 1$	$\hat{c}_s$	.988	.166	.830	.297	.696	.406	.584	-77.560
	$\widehat{SD}(c_s)$	(.014)	(.025)	(.025)	(.039)	(.039)	(.049)	(.049)	
$k = 2$	$\hat{c}_s$	.989	.100	.895	.285	.709	.501	.488	-71.942
	$\widehat{SD}(c_s)$	(.015)	(.022)	(.023)	(.038)	(.038)	(.056)	(.056)	
$k = 3$	$\hat{c}_s$	.988	.071	.923	.264	.730	.555	.431	-70.132
	$\widehat{SD}(c_s)$	(.015)	(.022)	(.033)	(.037)	(.037)	(.060)	(.060)	

Table 6: Gibbs Approximation to the CPO for the Prior with $\alpha_j = 1$ and $\gamma_j = .01$ for all j

	s	1-50	51-52	53-100	101-110	111-150	151-180	181-200	$\ln(\prod \hat{c}_s)$
$k=1$	$\hat{c}_s$	.988	.167	.829	.298	.695	.407	.582	-77.623
	$\widehat{SD}(c_s)$	(.015)	(.025)	(.025)	(.039)	(.039)	(.049)	(.049)	
$k=2$	$\hat{c}_s$	.989	.100	.894	.286	.708	.501	.487	-71.997
	$\widehat{SD}(c_s)$	(.015)	(.022)	(.023)	(.038)	(.038)	(.056)	(.056)	
$k=3$	$\hat{c}_s$	.988	.073	.922	.264	.730	.556	.431	-70.122
	$\widehat{SD}(c_s)$	(.015)	(.023)	(.024)	(.037)	(.037)	(.061)	(.060)	

Table 7: Gibbs Approximation to the Pseudo-Marginal Likelihood $\ln C(k)$

k	1	2	3	4	5	6	7	8	9
$\alpha_j = 2$ & $\gamma_j = 20$	-78.242	-72.743	-70.863	-70.358	-70.368	-70.784	-71.334	-71.959	-72.912
$\alpha_j = \gamma_j = 1$	-77.560	-71.942	-70.132	-69.856	-69.441	-69.764	-70.065	-70.058	-70.423
$\alpha_j = 1$ & $\gamma_j = .01$	-77.623	-71.997	-70.122	-69.798	-69.743	-69.603	-69.864	-70.044	-70.425

Table 8: Gibbs Approximation to the Bayes Estimates of β_1

k	1	2	3	4	5	6	7	8	9
$\alpha_j = 2$ & $\gamma_j = 20$	.161	.037	.030	.027	.026	.026	.025	.024	.024
$\alpha_j = \gamma_j = 1$	.175	.029	.025	.024	.024	.024	.023	.024	.025
$\alpha_j = 1$ & $\gamma_j = .01$	.175	.032	.025	.025	.024	.024	.024	.024	.025

6. Concluding Comments

We have illustrated the usefulness of the Gibbs sampling method in the linearized multi-stage model. The results can also be extended to the hierarchical Bayesian setup where the hyperparameters α_j and γ_j are random with appropriate distributions. The implementation of the Gibbs method can be carried out by methods similar to those in Gelfand and Smith (1990). When there is no conjugacy structure involved, the Metropolis algorithm (Metropolis, et al. 1953) can be used for sampling the variates. The results can also be extended to different prior distributions; for example, instead of gamma distribution, we can assume that each β_j has a truncated normal distribution, because $\beta_j > 0$. Then we can specify that the mean and variance of the normal distribution have normal and inverse gamma distributions. The Gibbs method can be implemented as in Section 2.

References

[1] Albert, J., and Chib, S. (1991), "Bayesian Analysis of Binary and Polychotomous Response Data," preprint.

[2] Armitage, P., and Doll, R. (1960), "Stochastic Models for Carcinogenesis," *Proceedings of the Fourth Berkeley Symposium,* ed. J. Neyman, 4, 19-38.

[3] Box, G. (1980), "Sampling and Bayes' Inference in Scientific Modeling and Robustness" (with discussion), *Journal of the Royal Statistical Society,* Ser. A, 143, 382-430.

[4] Crump, K.S. (1979), "Dose Response Problems in Carcinogenesis," *Biometrics*, 35, 157-167.

[5] Crump, K.S., Hoel, D.G., Langley, C.H., and Peto, R. (1976), "Fundamental Carcinogenic Processes and Their Implications for Low Dose Risk Assessment," *Cancer Research*, 36, 2973-2979.

[6] Geisser, S., and Eddy, W. (1979), "A Predictive Approach to Model Selection," *Journal of the American Statistical Association*, 74, 153-160.

[7] Gelfand, A.E., Dey, D.K., and Chang, H. (1992), "Model Determination Using Predictive Distributions with Implementation via Sampling-Based Methods" (with discussion), in *Bayesian Statistics 4*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford: Oxford University Press, pp. 147-169.

[8] Gelfand, A.E., and Kuo, L. (1991), "Nonparametric Bayesian Bioassay Including Ordered Polytomous Response," *Biometrika*, 78, 657-666.

[9] Gelfand, A.E., and Smith, A.F.M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398-409.

[10] Gelman, A.E., and Rubin D. (1992), "Inference from Iterative Simulation Using Multiple Sequences," preprint.

[11] Geman, S., and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

[12] Howe, R. B., Crump, K. S., and Van Landingham, C. (1986), "GLOBAL 86: A Computer Program to Extrapolate Quantal Animal Toxicity Data to Low Doses," unpublished EPA software documentation.

[13] Krewski, D., and Van Ryzin, J. (1981), "Dose Response Models for Quantal Response Toxicity Data," in *Statistics and Related Topics*, eds. M. Csörgő, D. A. Dawson, J. N. K. Rao, and A. K. Md. E. Saleh, Amsterdam: North Holland, pp. 210-231.

[14] McCullagh, P., and Nelder, J.A. (1989), *Generalized Linear Models*, (2nd ed.), London: Chapman and Hall.

[15] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equation of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21, 1087-1092.

[16] Nagaraj, N. K. (1989), "Bayesian Estimation in Dose-Response Models," preprint.

[17] Petrasovitas, A., and Cornell, R. (1975), "Approximations to the Bayes Estimate for a Quantal Assay with Simple Exponential Tolerance Distribution," *Communications in Statistics*, 4(9), 851-862.

[18] Pettit, L. I., and Young, K. D. S. (1990), "Measuring the Effects of Observations on Bayes Factors," *Biometrika*, 77, 455-466.

[19] Ramgopal, P., Laud, P. W., and Smith, A. F. M. (1992), "Nonparametric Bayesian Bioassay with Prior Constraints on the Shape of the Potency Curve," preprint.

[20] Stone, M. (1974), "Cross-Validatory Choice and Assessment of Statistical Predictions," *Journal of the Royal Statistical Society*, Ser. B,

111-147.

[21] Tanner, M., and Wong, W. (1987), "The Calculation of Posterior Distributions by Data Augmentation" (with discussion), *Journal of the American Statistical Association,* 81, 82-86.

[22] Zellner, A., and Rossi, P.E. (1984), "Bayesian Analysis of Dichotomous Quantal Response Models," *Journal of Econometrics,* 25, 365-393.